

**Claims:**

5 1) A natural language information extraction system for deriving information from a textual representation of a sentence, the sentence having a plurality of words, said system comprising:

10 a) an input for receiving data elements indicative of the textual representation of the sentence;

b) a processing unit coupled to said input, said processing unit being operative for processing the textual representation of the sentence to derive an information record on the basis of a set of information extraction rules, the information record being indicative of a semantic representation of at least part of the sentence;

c) an output for releasing the information record.

20 2) A natural language information extraction system as defined in claim 1, wherein said processing unit comprises:

i. a morphological analyser for assigning to each word in the sentence a most likely morphological tag;

25 ii. a syntactic processor coupled to the morphological analyser, said syntactic processor being operative for generating a parse tree group including a plurality of parse trees, each parse tree including a word of the sentence, at least one parse tree including at least two words of the sentence, said at least one parse tree including a dependency data element describing a

syntactic relationship between the at least two words of the sentence;

iii. an information extraction unit for processing the plurality of parse trees to generate an information record on a basis of the set of information extraction rules.

3) A natural language information extraction system as defined in claim 2, wherein each information extraction rule in said set of information extraction rules includes a data element indicative of a parse tree pattern, said information extraction unit being operative to process the plurality of parse trees to extract a certain parse tree of said parse tree group, the certain parse tree matching the parse tree pattern.

4) A natural language information extraction system as defined in claim 2, wherein said morphological analyser is operative for:

a) processing the textual representation of the sentence to assign a respective set of morphological tags to each word in the plurality of words;

b) assigning to each word an ambiguity class at least in part on the basis of the respective set of morphological tags;

c) identifying a most likely morphological tag to be associated to each word on the basis of a contextual rule, the contextual rule being associated to the respective ambiguity class assigned at step b).

5) A natural language information extraction system as defined in claim 2, wherein the syntactic processor is operative for:

5 i. generating a parse tree for each word in the sentence and adding each generated parse tree to a parse tree group;

10 ii. generating a new parse tree on the basis of binary dependency rules applied to a given parse tree in the parse tree group, the new parse tree resulting from a combination of the given parse tree and another parse tree from the parse tree group;

iii. adding the new parse tree to the parse tree group.

15 6) A method for deriving information from a textual representation of a sentence, the sentence having a plurality of words, said method comprising:

a) receiving data elements indicative of the textual representation of the sentence;

20 b) processing the textual representation of the sentence to derive an information record on the basis of a set of information extraction rules, the information record being indicative of a semantic representation of at least part of the sentence.

25

7) A computer readable medium comprising a program element suitable for execution by a computing apparatus for deriving information from a textual representation of a sentence, the sentence having a plurality of words, said  
30 computing apparatus comprising:

a) a processor, said program element when executing on said processor being operative for:

i. receiving data elements indicative of the textual representation of the sentence;

5 ii. processing the textual representation of the sentence to derive an information record on the basis of a set of information extraction rules, the information record being indicative of a semantic representation of at least part of the sentence;

10

iii. releasing the information record.

8) An apparatus for assigning a morphological tag to a given word in a sentence, the sentence including a set of words, said apparatus comprising:

15

a) an input for receiving a textual representation of the sentence;

b) a processing unit coupled to said input, said processing unit being operative for:

20

i. processing the textual representation of the sentence to assign a respective set of morphological tags to each word in the set of words including the given word;

25

ii. assigning to the given word an ambiguity class at least in part on the basis of a set of morphological tags associated to the given word;

30

iii. identifying a most likely morphological tag to be associated to the given word on the basis of a contextual rule, the contextual rule being associated to the ambiguity class assigned at step ii;

c) an output for releasing a signal indicative of the most likely morphological tag associated to the given word.

9) An apparatus as defined in claim 8, wherein an ambiguity  
5 class is associated to a collection of words.

10) An apparatus as defined in claim 8, wherein the  
contextual rule is indicative of a data structure  
associating a certain precursor context and a certain  
10 successor context in which a word occurs to a data element  
selected from the set consisting of a morphological tag,  
an ambiguity class, and a word.

11) An apparatus as defined in claim 10, wherein identifying  
15 a most likely morphological tag to be associated to the  
given word includes the application of a plurality of  
contextual rules.

12) An apparatus as defined in claim 11, wherein identifying  
20 a most likely morphological tag to be associated to the  
given word includes the recursive application of  
contextual rules.

13) An apparatus as defined in claim 8, wherein the  
25 contextual rule is indicative of a data structure  
associating a certain precursor context in which a word  
occurs to a data element selected from the set consisting  
of a morphological tag, an ambiguity class, and a word.

14) An apparatus as defined in claim 8, wherein the contextual rule is indicative of a data structure associating a certain successor context in which a word occurs to a data element selected from the set consisting  
5 of a morphological tag, an ambiguity class, and a word..

15) A method for assigning a morphological tag to a given word in a sentence, the sentence including a set of words, said method comprising:

- 10 a) receiving a textual representation of the sentence;  
b) processing the textual representation of the sentence to assign a respective set of morphological tags to each word in the set of words;  
c) assigning to the given word an ambiguity class at least  
15 in part on the basis of a set of morphological tags associated to the given word;  
d) identifying a most likely morphological tag to be associated to the given word on the basis of a contextual rule, the contextual rule being associated to  
20 the ambiguity class assigned at step c);  
e) releasing a signal indicative of the most likely morphological tag associated to the given word.

16) A computer readable medium comprising a program element  
25 suitable for execution by a computing apparatus for assigning a morphological tag to a given word in a sentence, the sentence including a set of words, said computing apparatus comprising:

- a) a processor, said program element when executing on said  
30 processor being operative for:

- i. receiving a textual representation of the sentence;
- ii. processing the textual representation of the sentence to assign a respective set of morphological tags to each word in the set of words;
- 5    iii. assigning to the given word an ambiguity class at least in part on the basis of a set of morphological tags associated to the given word;
- iv. identifying a most likely morphological tag to be associated to the given word on the basis of a
- 10    contextual rule, the contextual rule being associated to the ambiguity class assigned at step iii;
- v. releasing a signal indicative of the most likely morphological tag associated to the given word.

- 15 17) An apparatus for parsing a textual representation of a sentence to derive a parse tree group including a plurality of parse trees, the sentence including a plurality of words, the apparatus comprising:
  - a) an input for receiving data elements indicative of the
  - 20    textual representation of the sentence;
  - b) a processing unit for processing the data elements indicative of the sentence to generate a parse tree group, said processing unit being operative for:
    - i. generating a parse tree for each word in the
    - 25    sentence and adding each generated parse tree to the parse tree group;
    - ii. generating a new parse tree on the basis of binary dependency rules applied to a given parse tree in the parse tree group, the new parse tree resulting from a

combination of the given parse tree and another parse tree from the parse tree group;

iii. adding the new parse tree to the parse tree group;

c) an output for releasing a signal indicative of the parse tree group.

18) An apparatus as defined in claim 17, wherein each parse tree in the parse tree group includes a root node, the root node being associated to a word in the sentence.

19) An apparatus as defined in claim 18, wherein said apparatus is further operative for:

a) extracting a given parse tree from the parse tree group, the given parse tree having a root node associated to a given word;

b) processing a second parse tree, the second parse tree having a root node associated to a word that is a precursor to the given word to derive a dependency data element resulting from a combination of the given parse tree and the second parse tree;

c) combining the given parse tree and the second parse tree at least in part on a basis of the dependency data element to generate the new parse tree.

20) An apparatus as defined in claim 19, wherein the given parse tree covers a first range of consecutive words in the sentence, the second parse tree covering a second range of consecutive words in the sentence, the second range of consecutive words being the immediate precursor of the first range of consecutive words in the sentence.



21) An apparatus as defined in claim 18, wherein said apparatus is further operative for:

5 a) extracting a given parse tree from the parse tree group, the given parse tree having a root node associated to a given word;

10 b) processing a second parse tree, the second parse tree having a root node associated to a word that is a successor to the given word to derive a dependency data element resulting from a combination of the given parse tree and the second parse tree;

15 c) combining the given parse tree and the second parse tree at least in part on the basis of the dependency data element to generate a new parse tree, the new parse tree forming the new parse tree.

22) An apparatus as defined in claim 17, wherein said apparatus is further operative for:

20 a) searching the parse tree group for a parse tree that matches the new parse tree;

b) adding the new parse tree to the parse tree group if no matching parse tree is found at step a).

23) A method for parsing a textual representation of a sentence to derive a parse tree group including a plurality of parse trees, the sentence including a plurality of words, said method comprising:

a) receiving data elements indicative of the sentence;

30 b) processing the data elements indicative of the sentence to generate a parse tree group by:

- i. generating a parse tree for each word in the sentence and adding each generated parse tree to the parse tree group;
- ii. generating a new parse tree on the basis of binary dependency rules applied to a given parse tree in the parse tree group, the new parse tree resulting from a combination of the given parse tree and another parse tree from the parse tree group;
- iii. adding the new parse tree to the parse tree group;

24) A computer readable medium comprising a program element suitable for execution by a computing apparatus for parsing a textual representation of a sentence to derive a parse tree group including a plurality of parse trees, the sentence including a plurality of words, said computing apparatus comprising:

a) a processor, said program element when executing on said processor being operative for:

- i. receiving data elements indicative of the sentence;
- ii. generating a parse tree for each word in the sentence and adding each generated parse tree to a parse tree group;
- iii. generating a new parse tree on the basis of binary dependency rules applied to a given parse tree in the parse tree group, the new parse tree resulting from a combination of the given parse tree and another parse tree from the parse tree group;
- iv. adding the new parse tree to the parse tree group;

v. releasing a signal indicative of the parse tree group.

25) An apparatus for creating lexical frames from a parse  
5 tree group, the parse tree group being derived from a  
sentence including a plurality of words, said apparatus  
comprising:

a) an input for receiving a parse tree group including a  
plurality of parse trees, at least some parse trees  
10 including at least two words and a data element  
indicative of the syntactic dependencies between the at  
least two words;

b) a processing unit for processing the parse tree group to  
generate a plurality of lexical frames, each lexical  
15 frame being associated to a respective word in the  
sentence, a certain lexical frame being associated to a  
certain word in the sentence and comprising a list of  
words of the sentence other than the certain word, each  
word in the list of words being associated to a  
20 dependency data element indicative of the syntactic  
relationship of each word in the list of words with the  
certain word;

c) an output for releasing the plurality of lexical frames.

25 26) An apparatus for assigning semantic types to a noun  
phrase, the noun phrase including a set of words, said  
apparatus comprising:

a) an input for receiving a data element indicative of a  
noun phrase;

30 b) a processing unit operative for:

- i. assigning to each word in the noun phrase a semantic type on the basis of entries in a semantics dictionary to derive a sequence of semantic typed words, the semantics dictionary including a plurality of entries, each entry being indicative of a word associated to at least one semantic type;
  - ii. processing the sequence of semantic typed words on the basis of a set of semantic rules to derive a semantic type associated to the noun phrase, the set of semantic rules describing how to attach a semantic type to a given noun phrase;
  - c) an output for releasing a signal indicative of the semantic type associated to the noun phrase.
- 15 27) An apparatus as defined in claim 26, wherein the set of semantic rules includes at least one rule based on the presence of a semantic type associated to a word in the noun phrase.
- 20 28) An apparatus as defined in claim 26, wherein the set of semantic rules includes at least one rule based on the capitalization of words in the noun phrase.
- 25 29) An apparatus as defined in claim 26, wherein the set of semantic rules includes at least one rule based on the presence of specific words in the noun phrase.
- 30) An apparatus as defined in claim 26, wherein said noun phrase is a first noun phrase and the semantic type of the

noun phrase is a first semantic type, said processing unit being further operative for:

- a) processing a second noun phrase to derive a second semantic type associated to the second noun phrase;
- 5 b) combining the first noun phrase with the second noun phrase to derive a compound noun phrase on the basis of joining rules;
- c) assigning a third semantic type to the compound noun phrase;
- 10 d) releasing at the output a signal indicative of the compound noun phrase associated with the third semantic type.

31) An apparatus as defined in claim 30, wherein said  
15 joining rules are based on either one of the first semantic type and the second semantic type.

32) An apparatus as defined in claim 30, wherein said  
20 joining rules are based on the semantic types of words in either one of the first noun phrase and the second noun phrase.

33) An apparatus as defined in claim 30, wherein said  
25 joining rules are based on specific words in either one of the first noun phrase and the second noun phrase.

34) An apparatus as defined in claim 30, said processing  
unit is further operative for processing the compound noun  
phrase on the basis of the occurrence of incompatible  
30 semantic types within the compound noun phrase.

35) A method for assigning semantic types to a noun phrase, the noun phrase including a plurality of words, said method comprising:

- 5 a) providing a semantics dictionary including a plurality of entries, each entry being indicative of a word associated to a semantic type;
- b) providing a set of rules that describes how to attach a semantic type to a noun phrase;
- 10 c) receiving a data element indicative of a certain noun phrase;
- d) assigning to each word in the certain noun phrase a semantic type on the basis of entries in the semantics dictionary to derive a sequence of semantic typed words;
- 15 e) processing the sequence of semantic typed words on the basis of the set of rules to derive a certain semantic type associated to the certain noun phrase;
- f) releasing a signal indicative of the certain semantic type associated to the certain noun phrase.

20 36) A computer readable medium comprising a program element suitable for execution by a computing apparatus for assigning semantic types to a noun phrase, the noun phrase including a plurality of words, said computing apparatus comprising:

- 25 a) a processor, said program element when executing on said processor being operative for:
  - i. providing a semantics dictionary including a plurality of entries, each entry being indicative of
  - 30 a word associated to a semantic type;

- ii. providing a set of rules that describes how to attach a semantic type to a noun phrase;
  - iii. receiving a data element indicative of a certain noun phrase;
  - 5 iv. assigning to each word in the certain noun phrase a semantic type on the basis of entries in the semantics dictionary to derive a sequence of semantic typed words;
  - 10 v. processing the sequence of semantic typed words on the basis of the set of rules to derive a certain semantic type associated to the certain noun phrase;
  - 15 vi. releasing a signal indicative of the certain semantic type associated to the certain noun phrase.
- 37) A natural language information extraction system for deriving information from a textual representation of a sentence, the sentence having a plurality of words, said system comprising:
- 20 a) means for receiving data elements indicative of the textual representation of the sentence;
  - b) means for processing the textual representation of the sentence to derive an information record on the basis of a set of information extraction rules, the information record being indicative of a semantic representation of  
25 at least part of the sentence;
  - c) means for releasing the information record.